DOCUMENT RESUME

ED 377 248 TM 022 519

AUTHOR Griffin, Marlynn M.; And Others
TITLE Learning through Testing: An Investigation of Cooperative Assessment.
PUB DATE Apr 94
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Covariance; *Cooperative Learning; *Educational Assessment; Educational Psychology; Error of Measurement; Higher Education; *Learning; Outcomes of Education; Pretests Posttests; Research Methodology; *Student Attitudes; Test Format; *Test Use; *Undergraduate Students

ABSTRACT

Among the plethora of cooperative learning studies, several investigations of cooperative or group assessment (CA) have appeared. In this investigation, the methodology of R. R. McCown (1992) is refined to examine CA in a classroom setting, primarily examining achievement outcomes and student perceptions of the CA process. Fifty-six undergraduates in educational psychology participated in a nonequivalent control group design. It was hypothesized that students taking a test using a CA procedure would perform significantly better on a posttest of course concepts than would students completing the test in traditional format. The CA group completed examinations individually and in groups, while the traditional group (TA) took examinations individually. Analysis of covariance did not support the hypothesis. However, student reactions to the CA process were overwhelmingly positive. An appendix gives means, modes, and standard deviations for the group assessment survey. (Contains 13 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

MARLYNN M. GRIFFIN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 377 248

Learning Through Testing: An Investigation of Cooperative Assessment

Marlynn M. Griffin, Georgia Southern University
R.R. McCown, Duquesne University
Gwendolyn P. Quinn, Florida State University
Marcy P. Driscoll, Florida State University

Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 4-8, 1994.

TM022519

BEST COPY AVAILABLE

Learning Through Testing: An Investigation of Cooperative Assessment

Marlynn M. Griffin, Georgia Southern University
R.R. McCown, Duquesne University
Gwendolyn P. Quinn, Florida State University
Marcy P. Driscoll, Florida State University

Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 4-8, 1994.

Abstract

Amongst the plethora of cooperative learning studies, several investigations of collaborative, cooperative, or group assessment have appeared. These studies have investigated cooperative assessment in laboratory conditions (Lambiotte, et.al., 1987), and in classroom settings (Bilsky-Torna, 1993; Webb, 1993; McCown, 1992), and have examined effects of cooperative assessment on learning (Lambiotte, et.al., 1987; Bilsky-Torna, 1993; Webb, 1993; McCown, 1992) and group process on the assessment process (Webb, 1993). In this investigation, we have refined McCown's methodology and examined cooperative assessment in a classroom setting, primarily examining achievement outcomes and student perceptions of the cooperative assessment process.

Fifty-six undergraduate educational psychology students participated in the study. A non-equivalent control group design was utilized, and statistical analysis of the groups indicated that there were no statistical differences between the groups on grade point average and a pretest of educational psychology concepts. It was hypothesized that students taking tests using a cooperative assessment procedure would perform significantly better on a posttest of educational psychology course concepts than would students completing tests in a traditional format. The cooperative assessment group (CA) completed exams individually and in groups. Student exam grades were a combination of individual and group scores. The traditional assessment group (TA) took exams individually. Analysis of covariance indicated that there were no significant differences between the groups on a posttest of educational psychology concepts, thus the hypothesis was not supported. Student reactions to the cooperative assessment process were overwhelmingly positive. These data indicated that students enjoyed taking tests in groups, felt they learned more through this process, and that the CA procedures decreased test anxiety.

A growing body of research indicates that cooperative learning techniques enhance student achievement, self-esteem, and attitudes toward school at all grade levels, including in college classrooms (Johnson, Johnson, & Smith, 1990; Slavin, 1991). One explanation for these effects is that cooperative learning techniques augment the extent to which content is actively processed by students (Johnson et al., 1990). Cooperative techniques require students "to explain what they are learning to each other, learn each other's point of view, give and receive help from classmates, and help each other dig below the superficial level of understanding the material they are learning" (Johnson, et al., 1990, p. 11). While such "digging" and sharing of viewpoints can and does occur during in-class practice activities which take place as part of instruction, it is also likely that cooperative learning groups will "negotiate" under test conditions. Many believe that this negotiation of understanding is essential for knowledge construction (Duffy & Bednar, 1991; Kember & Murphy, 1990; Vygotsky, 1978). Brown (1989) argued that learning is "about the making of meaning, not just the receiving of it. Thoughtfulness is a constructive, not a passive, undertaking" (p. 32).

Lambiotte et al. (1987) found, under laboratory conditions, that cooperative assessment yielded increased learning. McCown (1992) reported similar findings in an unpublished pilot study of cooperative assessment. At least two other studies have investigated group or collaborative assessment on classroom, non-standardized tasks. Bilsky-Torna (1993) implemented group quizzes in a 10th grade English class. In this study, students took quizzes first as part of a group and then later took individual exams. All students in a group received the same score for the group quiz, but test scores were based on individual performance only. She found that the group quizzes were beneficial in several areas, especially to weaker students. The main advantages to group grades were: (1) increased motivation, resulting in a decrease of behavior problems; (2) teacher freedom (That is, with fewer papers to grade, the teacher is more free to provide extensive feedback on performance); (3) increased communication between teachers and students during group work; and (4) less cheating on quizzes. The main disadvantages which arose from this study included: (1) inability of some classes to easily form into groups; (2) instability of some groups; (3) possible difficulty in returning to traditional teaching methods; and (4) a noisy classroom. Other findings discussed in this study include: (1) weaker students riding on the coattails of the stronger students on group quizzes; (2) some "stifling" of stronger students when forced to work in groups; and (3) in several groups, the results of the group exam were higher than the highest individual score.

In another study, Webb (1993) investigated the correlation between achievement scores from small-group and individual assessment contexts. Group processes, as well as outcome data, were also examined. In this study, students solved mathematics operations on decimal numbers in

collaborative small groups for a 50-minute class period. Two weeks later, following a review session, students examined a similar problem on their own, without collaborating with other students. Student ability and behavior within the group setting both predicted performance on the individual assessment. Student performance in the cooperative group context was uniformly high and did not predict performance in the individual context, as individual performance was often lower than performance in the group setting.

Proponents of authentic assessment have called for assessment to be more than an end product. Rather, assessment should be part of the learning process (c.f., Shavelson, Baxter, & Pine, 1992; Shepard, 1989; Wiggins, 1989). Most educators have had the experience of returning a test to the class and attempting to help students learn from their mistakes, only to find that many of the students are far less interested in learning from their mistakes than they are in trying to rationalize their incorrect responses in the hopes of earning extra points. Cooperative assessment offers students an opportunity to see a test twice and to negotiate responses with peers on the second administration. This negotiation can be a valuable learning tool and can encourage students to think about what they have learned and are learning as they discuss the assessment with their peers.

This study investigated the effects of group assessment on individual learning in an Educational Psychology class for pre-service teachers. The design differs from that utilized by Lambiotte (1987) in that it occurs in a classroom context rather than a laboratory context, and is thus more ecologically valid. It also differs from those utilized by Webb (1993) and Bilsky-Torna (1993) in that students take tests first individually and then in groups to ensure individual accountability as well as group rewards. McCown (1992) found positive effects on student grades and student attitudes using this design.

Method

Design and Subjects: 56 students enrolled in two sections of Educational Psychology participated in the study. All students were education majors with a variety of areas of emphasis (early childhood, middle grades, physical education, etc.). There were 6 males and 29 females in the treatment group, and 5 males and 16 females in the control group. 5 students in the experimental and 4 students in the control group were Black, while the remainder were White.

A non-equivalent control group design was utilized. Pretest scores and self-reported GPAs were obtained and analyzed to determine if groups differed significantly in these areas. The groups were not significantly different (alpha = .05) on either pretest scores $F(1, 55)$, $p=.42$ or GPA, $F(1, 55)$, $p=.58$.

The independent variable was testing condition, in which participants either took tests first individually and then as part of a group (cooperative assessment or CA) or completed tests individually and then reviewed the exam with the instructor of the course (individual assessment or

IA). These reviews primarily consisted of providing the correct answers to the test items, but also included explanations of misconceptions if students requested. The exam review was incorporated as part of the IA group procedures to control for the effects of differential number of exposures to the assessments by the two groups.

Students in the treatment group could add up to ten points to their individual grade if performance on the group assessment exceeded their individual performance, but would not lose any points if individual performance surpassed group performance. Cooperative scores (the student's combined individual and group assessment scores) could not exceed group scores. For example, Group A obtained a 96 on their group exam. Student 1 obtained an 82 on her individual exam, and is thus eligible for all 10 group points, bring her cooperative assessment score to 92. Student 2 scored 88 on the individual exam, and is thus eligible for only 8 group points so that she does not exceed the group score, giving her a 96 on her exam. A score of 97 was obtained by Student 3 on the individual assessment, so she does not receive any group points and keeps her grade of 97. In this way, both individual accountability and group incentives were established. Since students could earn add to 10 points to their individual score, they had to prepare for the exam in order to do well. By allowing students to earn extra points to add to their individual grades, incentive to work together as part of a group was also established as it was in everyone's best interest to obtain the highest group score possible to maximize the likelihood that one would earn group points and thus increase their exam grade.

The dependent variable was performance on a posttest of educational psychology course content. Several sources of additional data were also examined and included: written comments and reactions to cooperative assessment from students following the first exam, survey data from the CA group evaluating the cooperative assessment strategy, and follow up interviews with a representative from each of the cooperative groups in the treatment condition.

Materials

Pretest. The pretest consisted of 30 multiple choice items, 10 from each of the first three units of instruction for an introductory educational psychology course. The topics addressed on the pretest were operant conditioning and information processing (unit 1), observational learning, motivation, and outcome decisions (unit 2), instructional models, instructional tactics, and classroom management (unit 3). The items on the pretest were primarily application type multiple choice items, and all were matched to unit objectives to ensure content validity.

Posttest. The posttest was comprised of the same 30 items which appeared on the pretest and addressed the **first** three units of instruction. The posttest was administered in three sections, 10 questions on each of the **last** three unit exams. Thus, the last three exams were partially cumulative. For example, the exam for unit 2 consisted of 30 items from the chapters for unit 2 and 10 items from the chapters discussed in unit 1. The number correct on the 10 cumulative items

from each of the last three exams was added together to form the posttest score. This partially cumulative approach to testing was designed to provide a measure of the testing condition effects on individual learning. Using this strategy, students completed test items from unit one by themselves on the individual portion of the test. Next, the cooperative assessment condition discussed the same items on unit 1 in their groups, while the instructor went over the test with students from the control condition. Then, on the second unit test, students completed the 10 posttest items from unit 1 (as well as the 30 items from unit 2) on an individual basis prior to discussing them with their groups or reviewing the test with the instructor. The posttest scores were taken from the **first** administration of the unit 2, 3, and 4 exam. Thus, although students in the cooperative assessment group did obtain higher grades on their tests (as a result of combining individual and group scores), they did not necessarily obtain higher scores on the posttest just by virtue of taking the test twice.

Group Assessment Survey. A 12 item survey was designed to tap student opinions and perceptions of the cooperative assessment process. The first items examined each student's perceptions of their own and the other group members' participation in the cooperative assessment process, and will not be discussed in this paper. The remaining nine items examined group process and student perceptions of the testing situation (see Appendix A).

Procedures

On the first day of the quarter, all students completed the 30 item pretest of educational psychology concepts. Students also completed a questionnaire requesting demographic information which was used to assign students to heterogenous groups.

Students in both classes were assigned to cooperative groups of 4-5 members, which were mixed on the basis of gender, race, ability, and major. In terms of ability, each group consisted of a student with a high GPA (3.5-4.0), two students with average GPAs (2.5-3.49), and one student with a low GPA (below 2.49). Males and Black students were divided among the groups, and, to the extent possible after mixing the groups on all other variables, each group contained at least one student who was not an early childhood education major (the predominant major in both classes). Students worked in their groups to complete in-class non-graded activities designed to facilitate acquisition of course concepts. Each group also worked to cooperatively complete a series of article critique papers dealing with issues in educational psychology. These papers were graded and each member of the group received the same grade on each of the papers (4-5 papers, depending on the number of people in the group). In addition, the students in the CA group completed the second administration of each unit exam with their cooperative group.

With the exception of the difference in testing procedures, both sections of the course were taught in the same manner. Care was taken by the instructor to use similar examples, to complete the same activities, and to cover the same amount of material in both classes. There were

occasions, however, on which one group discussed something to a different depth or from a different angle than the other group. Although this does not allow for strict control of the teaching conditions, more effort to deter these differences was not made because the instructor was unwilling to interfere with the quality of the instruction in order to implement strict experimental controls. Class sessions were typically a mixture of lecture, discussion, generation of examples, examination of practical applications of course content, and group activities.

At the end of each of the four units, students completed a unit examination. Each exam was comprised of 40 multiple choice questions (primarily higher level items) and 2 essay items. Exams were administered in a two-hour time block. Both groups had 75 minutes in which to complete the individual examinations. (Students in the experimental group were allowed to come to class early to begin the exams so they could have 75 minutes to complete the individual portion and an hour to complete the group portion). Students were allowed to leave the room after completing the individual portion of the exam and were instructed to return at a time designated by the instructor. Upon the students' return, the instructor reviewed the exam with students in the control group.

Students in the experimental group worked together to complete the test a second time. Each group was permitted to submit only one set of responses per group, so students were forced to come to consensus on the responses to the test items. Exams were returned to students within two class periods. During the period in which exams were returned, the instructor placed an answer key on the overhead projector so students could check the accuracy of the machine scoring. Copies of the exam were not distributed to students, but were available in the instructor's office at any time the student wanted to come in and see which items they answered incorrectly. Very few students chose to pursue this option.

On the final examination date, following completion of the exam, students were informed that they had been part of an experiment throughout the quarter. They were briefly told of the purpose of the study and assured that individual data would be kept confidential. At this point, students were also told that they could request that their data be withdrawn from the study if they chose not to participate. Students could exercise this option at any time, so that if they were afraid of retribution from the instructor they could choose to withdraw after their grades had been submitted to the registrar. No student chose to withdraw his or her results from the data pool. Recall that the students in the cooperative assessment condition could earn up to 10 point above their individual score on each examination, if the group score exceeded an individual's score. Students in the control group were also informed that their grades would be "curved" so that the distribution of grades in their class would match that from the experimental group. This measure was taken so as not to penalize the grades of students who selected the "wrong" section of the

class. Students in the experimental group completed the Group Assessment Survey on the day of the final examination.

Early in the Winter quarter (approximately 6 weeks after the end of the Fall quarter, during which time the treatment took place), nine students from the cooperative assessment condition were asked to come in and participate in individual interviews with the investigator. One student from each of the nine groups was selected and in all cases but one the student with the highest GPA in the group participated. These students were intentionally selected because they would be expected to benefit the least from the cooperative assessment process since they were already successful students based on their own individual talents. Thus, if they reported positive results on their understanding gleaned from the assessment process, it would seem reasonable that less skilled students might also experience this increased understanding. These interviews were audiotape recorded (with the permission of the student). Each student was asked the same series of questions, but responses were explored in more depth as needed.

Results

Posttest

All data were analyzed at the alpha = .05 level. Posttest data were analyzed using ANCOVA, with pretest scores and GPA utilized as covariates. No significant differences were found between the groups on the posttest, $F(1, 52)$ $p=.54$. Only GPA was significant, $F(1,52)$ $p=.007$, indicating a positive relationship between GPA and posttest performance. Thus, the hypothesis of this study was not supported.

Reactions to Cooperative Assessment (following first exam)

Following the first exam, students were asked to write brief, anonymous reactions to the assessment process. They were told that the instructor was trying a new procedure and were encouraged to indicate any concerns or criticisms about the process so modifications could be made, if necessary. 32 of 36 students chose to comment about the process, and 29 of the responses were extremely positive. Of the remaining 3 comments, one concern was related to group process, not the activity of testing. The other two comment will be discussed in more detail in the next section of this paper.

Survey

Analysis of survey data also indicates that student perceptions toward cooperative assessment were quite positive. Mean scores, modes, and standard deviations for the last nine items on the Group Assessment Survey are presented in Appendix A (note that items 9 and 12 are scored in reverse). These results indicate that students tended to discuss test items with group members, prepared about the same amount for exams in this course (despite the fact they would receive extra points from group collaboration) as they did for exams in other courses, and felt that

taking tests in groups was somewhat beneficial to their grades. Students also believed that they were learning more about the course content by discussing the exams with their groups.

Interview

The last source of data to be discussed, interview data, was compatible with the outcomes of the survey. Most students interviewed reported satisfaction with the cooperative assessment procedure, indicating that they felt they had learned more in this class than in comparable classes in which students take tests only on an individual basis. They also reported that they enjoyed working in groups and had few conflicts, if any, with their group members.

In sum, then, the outcomes on the posttest do not support the hypothesis of this study. Student response to the cooperative assessment strategy, however, was overwhelmingly positive.

Discussion

Overview and Posttest Outcomes

This study examined the effect of cooperative assessment on the acquisition of educational psychology course content. Students in the cooperative assessment group (CA) took exams first on an individual basis and then in groups. Their exam scores were a combination of their individual and group scores. Students in the individual assessment group (IA) took tests individually and then received feedback on their performance as the instructor reviewed the test with the group immediately after they had completed it.

It was predicted that students in the CA group would outperform students in the IA group on a posttest of of educational psychology. This hypothesis was not supported. Certainly the first explanation to be considered regarding this finding is that cooperative assessment does not improve learning over an individual assessment format, as the results seem to indicate. Student perceptions, however, seem to contradict this finding. As will be discussed in greater detail, students reacted very positively to the treatment and felt that they learned more from this method of testing than from traditional methods. Still, student perceptions are not always accurate indicators of what is actually occurring.

A second possible explanation for the NSD findings of this study lies in the nature of the control group test taking procedure. The original intent behind going over the test with students immediately after completion of the test was to control for the number of exposures to the exam. Reviewing the test with the students was not expected to be a tremendous learning opportunity for the students, it was intended to provide feedback regarding correct responses. Many instructors do not review tests in detail (including at least one of the authors of this study!), and even fewer attempt to do so immediately after students complete the exam. (Any of you who have gone through a test in detail know the vulnerable feeling which results from this practice!) The students, however, seized this time as an opportunity to debate responses to items, to provide rationales for selecting the incorrect responses they selected, and to ask the instructor to clarify to a greater depth

confusing material. Thus, there was tremendous potential for learning during test review. Survey data from a follow up study on cooperative assessment are still being analyzed, but results seem to indicate that students do, in fact, believe that they learn much from these review sessions. Furthermore, students in both the original and follow up study have said that they wish more teachers would review exams immediately after completion as they learn more when they are still in the testing mode and the feedback is more meaningful. Thus, it is possible that the control group procedures led to outcomes which mask any gains which might be obtained through cooperative assessment. At the very least, results indicate that students seem to learn equally well from both methods though one is more teacher directed and the other more student directed. Thus, if one is concerned with eliminating the teacher as giver of knowledge (Kember & Murphy, 1990), the cooperative assessment procedure may be an effective means for accomplishing this goal.

Reactions to Cooperative Assessment (following first exam - Data from CA Group only)

Only 3 of the 32 students (N=36) who chose to offer feedback about the first cooperative assessment activity offered negative comments. The negative feelings behind one of these comments were directed more at the group process, however, than at the cooperative assessment process: "I think it's a really good idea; however, it didn't work well with my group (one person wanted everything done her way...really pushy). I'm going to say something next time, so maybe it will work."

The second comment came from a rather insightful student who wrote, "I think that taking the test is redundant and probably is really not necessary, unless you are doing this for some kind of research project that you are conducting." Unfortunately, this student did not remain anonymous. While his comment may have been insightful in terms of realizing he was participating in a study, he was not particularly insightful in terms of his study habits for the course. This student chose not to contribute to his group's discussion of items on the group assessment; rather, he chose to study for other classes or to discreetly read the student newspaper. He apparently did not gain anything from the group discussions, as by the end of the quarter he was the only student in two classes (N=55) to earn an F in the course. On the other hand, perhaps he made a valid point with his observation.

The third negative comment, however, addressed a very important concern that often arises when students complete work in groups: "As far as the group testing goes, I believe it was helpful (but I have a good cooperative group). That could be a problem for some. I studied for the test, but I know I slacked up because my group works hard." This "slacking up" is often a concern (and a quite valid one) for those who dare to assign group work. While this may have been a problem for a few students in the class, it did not appear to affect many students.

In contrast to the 3 concerns about cooperative assessment, 29 students were quite positive in their perceptions of the process. Several students expressed opinions in direct contrast to the last concern stated above:

> "I like the way the test was given. I know it will help me. Even if I missed only one answer, that could still be one answer someone else got right. I don't think people would be slack in studying, because, personally, I wouldn't want to look stupid in front of other people. The only way I would have a problem with this testing method is if our grade could go down. Considering this cannot happen, I think the testing method is great. Anyone who doesn't think so needs to think again or get a brain so they can think to start with."

Other reactions to cooperative assessment addressed the potential for learning from this process.

> "I liked taking the test twice. I felt I retained the information better this way. It was more than taking a test and writing the answers down. Discussing it with the group I will remember it longer."

> "I like the idea of the group test because if I am confused about a certain area of questions, I can discuss it with my group, better understand it, and fix my mistake. A lot of time it helps to talk through questions."

> "Before taking the group test, I didn't think it was such a great idea. Although after having taken it, I realize that it was. I was [able] to see why I missed various questions and I was also able to see why my group members chose their answers. I learned a lot!"

> "I like the way you are doing the tests with the cooperative learning. It is also a good review for the tests that we just took since some of it might be on the next test. It's also interesting to see how others think."

Still other students focused more on the effects of the cooperative assessment on their grades. Though not as prevalent as the comments regarding the beneficial effects on learning, a few of these comments are listed below.

> "I like the testing thing - not only does it help our grades, but it helps us learn the material better."

> "I like the test-procedure - taking it twice. I feel I have a better chance of getting a good grade."

While some (7) of the positive comments did address the beneficial effects of the group testing on their grades, many more focused on the opportunities for learning that this method offered. Thus, though the posttest scores did not support the hypothesis, students did, in fact, feel as if they were learning through the cooperative assessment process.

Survey (data from CA Group only)

The results of the survey administered at the end of the quarter were quite compatible with the comments made by students following the first test. All survey items used a 7-point Likert scale with higher numbers indicating more positive responses (see Appendix A). Three major

areas were emphasized in the survey questions: group testing process, study habits, and benefits of cooperative assessment.

Items 4, 9, and 10 addressed the processes through which the groups went as they completed tests together. Note that item 9 is reverse scored, so that the higher mean corresponds with positive responses on items 4 and 10. The survey indicates that students did tend to discuss the responses to test items more than they tended to rely on one person to provide a correct response. Thus, students were negotiating the responses to test items and attempting to reach a mutual understanding about the content. The mean scores for these items are not as high as they could be, but this result is perhaps caused by the wording of the survey items. Students were required to consider all items on the test as they responded to items 4 & 9, and on some items there was discussion of correct response, but for others, group consensus was the norm. Observations of the group testing phase by the instructor indicated that students did not discuss all items; rather, for those on which all group members had selected the same response on the individual exam, this response was chosen by consensus on the group exam. For those items on which group members selected different responses, however, lively debates often ensued with different members trying to convince others of the "rightness" of "their" response. So, for low consensus items, students did tend to negotiate responses.

Study habits were also addressed by the survey. It seems quite reasonable that students might study less for an exam if they knew they would have the opportunity to improve their grades through group testing than if they were in a traditional testing situation in which the grade they earn is solely the result of their efforts. This possibility was, in fact, what prompted us to allow the group exam to count only a maximum of 10 points toward a student's exam grade. If students can earn 10 points, there is incentive to work with the group to earn the highest grade possible. There is also incentive, however, to prepare for the individual test since one must do reasonably well in order for the group points to be of benefit. Survey results indicate that students studied about the same amount as they usually do, despite knowing they had a group exam to complete. Of more consequence to the students' test preparation was the difficulty level of the exams - students indicated that they prepared somewhat more than usual due to the difficulty of the exams. Perhaps the group exams had a mitigating effect on the extra study efforts caused by test difficulty. Several students did report in interviews and in informal conversation that they liked the group testing because it helped to alleviate test anxiety. Though not part of the research hypothesis, both findings, the alleviation of test anxiety and the maintenance of usual study habits, are important outcomes of this research.

Last, the survey examined the potential benefits of the cooperative assessment method. As in interviews, students focused less on the benefit of the testing to their course grade than on the benefits to their understanding of course content. Students did feel that the cooperative assessment

was beneficial to their grade, but they seemed to feel that cooperative assessment was more beneficial to their understanding of course content than to their grade. Note that items 8 and 11 (effect on understanding) have higher means and lower standard deviations than item 7 (effect on course grade). Question 12 is also somewhat related to the students' understanding of course content. This item (which is reverse scored) indicates that students were also concerned with receiving feedback from the group assessment, but were somewhat more concerned with the benefits of learning from their group. Overall, then, students believed that the cooperative assessment procedure did benefit their understanding of course material. This finding is another plus for using cooperative assessment - students believe they are learning through this method as they work in groups to complete testing tasks. Again, group discussion of exams takes the responsibility for dispensing knowledge to the students out of the hands of the teacher and places it into the hands and minds of the students, encouraging them to be responsible for their own acquisition and construction of knowledge.

Interview (data from CA Group only)

Interview data are also compatible with comments made by students following the first exam, as well as with the results of the survey data. The interviews focused primarily on issues related to group process, the testing process, and benefits of cooperative assessment. Examining group process was not intended to be a major goal of this study, but several incidents which occurred during the course of the quarter indicated that, at the very least, this area requires more attention in future studies. Webb (1989) has examined group process as it relates to learning in groups and relative to cooperative assessment (1993). We chose not to focus on these issues because we wanted to explore the outcomes of cooperative assessment without doing any group intervention.

Interview results indicate that, in future studies, group process and the testing process should be addressed. No major incidents occurred, but three groups reported frustration with a group member who did not do his or her "fair share" throughout the quarter. All of these groups informed the instructor of the course about these problems, but none chose to address it with the "offenders." At the very least, this indicates a need for teaching some sort of conflict resolution skills. Another area in which some intervention is needed (and was provided in a subsequent study of cooperative assessment) is the group assessment process itself. That is, what should students be doing as they complete the exams together? During this study, students were just told to complete the exams together. They were not given instructions or guidelines for interaction. As discussed above, in many instances students did not choose to discuss the test items on which they reached group consensus. Interview results further reveal that students, on occasion, did not choose to debate those items on which there was disagreement. In some groups, students would briefly discuss the problematic items then rely on one group member to provide the response or

would alternate the response selected on different non-consensus items. Future studies should provide training for students in group assessment and conflict resolution skills.

Interviewees were asked several questions related to the benefits of taking tests in groups. When asked if they learned anything by taking the tests twice, the students being interviewed had mixed responses. Three of the students felt that they didn't learn much, but were able to get a better angle on a few course topics by taking the tests together. The other 6 respondents indicated that they did feel they learned more taking the tests twice. This increased understanding was primarily due to the opportunity to discuss the test material within the groups. One student reported, "Yes [it was helpful to take tests twice], because if I missed something I could hear why. It also helped to explain the information to the other people." It is important to remember that although the interviewed students did not overwhelmingly report strong benefits to their understanding through the group exam, these students were also the students from each group with the highest GPA. Thus, these students were already skilled learners and would be expected to be less likely than less skilled learners to benefit from the group discussion. Students did report, however, that they believed that the others members of their group benefitted from the group testing process. It is interesting to note that the student who seemed least impressed with the benefits of the group testing to her understanding of course content also noted that if given the opportunity she would take another class which offered cooperative assessment. When asked why, she responded, "I didn't learn much from taking the tests in groups this time, but there is always the possibility that the group tests would help me understand the class better." Students reported that they felt they learned more taking cooperative assessments than they did in classes which did not offer this strategy.

Summary

Cooperative assessment did not lead to higher posttest scores than individual assessment followed by a teacher review of the exam. Besides the obvious explanation that cooperative assessment is not more effective for learning, it is possible that the control group's testing strategy was also beneficial in increasing understanding and masked the effects of the treatment.

Student perceptions of cooperative assessment were positive. Anecdotal, survey, and interview data indicate that students enjoyed the cooperative assessment process and believed they learned more by taking the tests twice and discussing their responses with other students. Several students also reported that this technique reduced test anxiety. Regardless of the NSD outcome for the hypothesis of the study, these affective outcomes are quite noteworthy. Improving student perceptions of the learning potential of exams and decreasing test anxiety are worthy goals for an assessment intervention.

This is only the beginning of a series of studies designed to investigate cooperative assessment. Future studies should examine this strategy with different types of tests in different

content areas with learners of varying ages. The cooperative assessment process itself should be scrutinized and refined. The effects of cooperative assessment on motivation, achievement, self-efficacy and student perceptions of learning should also be explored.

References

Bilsky-Torna, Z. (1993). "Son of S.T.A.D." - A comparison of group quiz and individual test scores. English Teachers Journal, 46, 35-43.

Brown, R. (1989). Testing and thoughtfulness. Educational Leadership, 46, 31-33.

Duffy, T.M. & Bednar, A.K. (1991). Attempting to come to grips with alternative perspectives. Educational Technology, 31, (September), 12-15.

Johnson, R.T., Johnson, D.W., & Smith, K.A. (1990). Cooperative learning: An active learning strategy for the college classroom. Baylor Educator, 15 (2), 11-16.

Kember, D. & Murphy, D. (1990). Alternative new directions for instructional design. Educational Technology, 30(8), 42-47.

Lambiotte, J.G., Dansereau, D.F., Rocklin, T.R., Fletcher, B., Hythecker, V.I., Larson, C.O., & O'Donnell, A.M. (9187). Cooperative learning and test taking: Transfer of skills. Contemporary Educational Psychology, 12, 52-61.

McCown, R.R. (1993, July). Personal communication.

Shavelson, R.J., Baxter, G.P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. Educational Researcher, 21(4), 22-27.

Shepard, L.A. (1989). Why we need better assessments. Educational Leadership, 46(April), 4-9.

Slavin, R.E. (1991). Synthesis of research on cooperative learning. Educational Leadership, 48(February), 71-82.

Vygotsky, L. (1978). Mind and society. Cambridge, MA: Harvard University Press.

Webb, N.M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. Educational Assessment, 1(2), 131-152.

Wiggins, G. (1989). Teaching to the (authentic) test. Educational Leadership, 46(April), 41-47.

Appendix A

Means, Modes, and Standard Deviations for the Group Assessment Survey

Item	Mean	Mode	SD
4. In general, our group discussed and debated the answers to most of the test items.	6.09	7	1.2
5. What effect did taking the group exam have on your study habits?*	4.83	4	0.95
6. What effect did the difficulty level of the exams have on your study habits?*	5.29	6	0.93
7. Taking the tests as part of a group was beneficial to my grade.	5.64	7	1.72
8. Taking the tests as part of a group was beneficial to my understanding of required course concepts.	6.4	7	0.85
9. In general, our group did not debate the responses to most items, but relied upon one or two members to provide the answers.	5.17	7	1.95**
10. The group as a whole worked together to complete all of the essay items.	5.49	6	1.42
11. As I completed the group exam, I gained a better understanding of content I missed when I took the test individually.	6.57	7	1.01
12. As I completed the group exam, I was more concerned with determining which items I answered correctly on my own than with learning from the group.	5.17	7	1.77**

All items used a 7-point Likert scale, with higher numbers indicating more positive responses.

*End points on the Likert scale for these items were: "I studied much less than usual" (1) to "I studied much more than usual" (7). All other items used "Strongly disagree" (1) to "Strongly agree" (7).

**These items were reverse scored. Reversed scores are reported in table.